

Polarization Effects in Generalized Molecular Interaction Potential: New Hamiltonian for Reactivity Studies and Mixed QM/MM Calculations

F. JAVIER LUQUE,¹ MODESTO OROZCO²

¹*Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Avda. Diagonals/n, Barcelona 08028, Spain*

²*Departament de Bioquímica, Facultat de Química, Universitat de Barcelona, Barcelona, Spain*

Received 9 May 1997; accepted 12 January 1998

ABSTRACT: A strategy to introduce multipole-induced polarization effects into quantum mechanical particles is presented. The strategy takes advantage of perturbation theory, and allows one to introduce polarization effects into the generalized molecular interaction potential (GMIP) previously defined by our group. The resulting parameter provides a complete and accurate picture of molecular interactions at a reduced computational cost. It is suggested that this polarized potential (GMIPp) can be used as an effective Hamiltonian for mixed QM/MM calculations. © 1998 John Wiley & Sons, Inc. *J Comput Chem* 19: 866–881, 1998

Introduction

Molecular interactions can be treated at three levels: (i) quantum mechanical (QM); (ii) classical (MM); and (iii) mixed (QM/MM). This latter level combines a QM representation of a small portion of the system (one or few molecules), while the rest of the system is treated classically.

Correspondence to: F. J. Luque

Contract/grant sponsor: Centro de Supercomputació de Catalunya

Contract/grant sponsor: Spanish DGICYT; contract/grant numbers: PB94-0940, PB93-0779

Mixed QM/MM algorithms are suitable for the study of chemical processes in large molecular systems, where only those chemical species directly involved in the process of interest need a QM treatment, while the rest of molecules are described using effective potentials derived from classical force fields. Thus, mixed methods are valuable for the representation of dilute solutions, chemical reactivity in condensed phases, or the study of biological macromolecules.^{1,2}

The effective Hamiltonian of the system in QM/MM methods consists of three terms [eq. (1)]: (i) the Hamiltonian for the QM molecule in the gas phase, which throughout this article will be denoted by *A*; (ii) the molecular mechanical energy

for the classical portion of the system, which will be indicated by B ; and (iii) the coupling term between QM and MM subsystems, with classical atoms being treated as charged Lennard–Jones particles. Accordingly, the coupling term consists of Coulombic interactions between classical atoms with nuclei and electrons of the QM molecule, and of van der Waals interactions with the nuclei of the QM particle:

$$\hat{H}_{\text{eff}} = \hat{H}_A^0 + \hat{H}_B + \hat{H}_{A-B} \quad (1)$$

The electrostatic coupling between QM and classical subsystems can be treated at three levels: (i) the charge distribution of A is not altered by the charges of B ; (ii) A is polarized by B , which changes the charge distribution of the QM molecule; and (iii) all the molecules are polarized by the rest of the system. In the first case, the electrostatic component of the interaction energy between QM and MM subsystems is determined as shown in eq. (2a), where Ψ^0 is the wave function of the unperturbed (nonpolarized) molecule A and H_{A-B}^{el} is the electrostatic interaction Hamiltonian [eq. (2b)]. In these equations, Z_A denotes the nuclear charge of the molecule: A ; Q_B is the partial charge of classical atoms B ; and r , R_A , and R_B stand for the position of electrons and nuclei of A , and of classical atoms of B , respectively.

$$E_{A-B}^{el} = \langle \Psi^0 | \hat{H}_{A-B}^{el} | \Psi^0 \rangle + \sum_A \sum_B \frac{Z_A Q_B}{|R_A - R_B|} \quad (2a)$$

$$\hat{H}_{A-B}^{el} = \sum_B \frac{Q_B}{|r - R_B|} \quad (2b)$$

The use of eq. (2a) describes the A – B electrostatic interaction more accurately than purely molecular mechanics calculations. Nevertheless, it does not account for the energy contribution arising upon polarization of the QM molecule by classical particles. This contribution can be very important when the QM subsystem is located in a very polar environment, such as dilute aqueous solutions.^{3–7} Inclusion of the $B \rightarrow A$ polarization can be achieved by using eq. (3), where the wave function of A is determined by solving the pseudo-Schrödinger given in eq. (4). The use of eqs. (3) and (4) implies that a full SCF process has to be performed to obtain the new wave function of A for each configuration of the classical particles:

$$E_{A-B}^{el} = \langle \Psi | \hat{H}_A^0 + \hat{H}_{A-B}^{el} | \Psi \rangle - \langle \Psi^0 | \hat{H}_A^0 | \Psi^0 \rangle \quad (3)$$

$$(\hat{H}_A^0 + \hat{H}_{A-B}^{el})\Psi = E\Psi \quad (4)$$

The $B \rightarrow A$ polarization energy is given by the difference between eqs. (3) and (2a), which yields eq. (5):

$$E_{\text{pol}} = \langle \Psi | \hat{H}_A^0 + \hat{H}_{A-B}^{el} | \Psi \rangle - \langle \Psi^0 | \hat{H}_A^0 + \hat{H}_{A-B}^{el} | \Psi^0 \rangle \quad (5)$$

The introduction of the mutual polarization between QM and classical particles is more difficult, and often less relevant for the total value of the A – B interaction energy. The $A \leftrightarrow B$ polarization is typically accounted for using approximate expressions derived from induced-dipole theories,⁸ which can also include the polarization between classical particles. Treatment of the $A \leftrightarrow B$ polarization implies the use of polarizable force fields, which are less popular than standard force fields, and the cost of the calculation is sensibly larger. In fact, very few fully polarized QM/MM calculations have been reported in the literature.^{9–13}

In the usual case of a QM solute surrounded by classical polar solvent molecules, QM/MM calculations are performed within molecular dynamics or Monte Carlo algorithms. The inclusion of polarization effects, even when using the simple eq. (5), requires a very large number of SCF calculations (in the range of millions for Monte Carlo simulations). This makes QM/MM calculations very expensive and justifies the use of low-level Hamiltonians (semiempirical, local density functional, etc.) to treat the QM particle.^{14–27}

In this article a new approach to define the coupling QM–MM interaction energy is presented. The strategy is based on the introduction of polarization effects into the generalized molecular interaction potential (GMIP), previously developed by our group.^{28–30} The polarization is accounted for following a perturbational treatment,³¹ which avoids the need to perform the SCF calculation for each configuration of the classical particles. The resulting parameter, called generalized molecular interaction potential with polarization (GMIPp), is useful as an effective Hamiltonian in mixed QM/MM calculations, and is also a major tool to describe the reactive characteristics of a QM molecule in front of different interacting molecules.

Theoretical Framework

A typical decomposition analysis of the interaction energy between two molecules considers two

main terms: (i) the SCF interaction energy; and (ii) the dispersion contribution. A further partitioning of the SCF term can be achieved following different schemes, like those developed by Bonaccorsi et al.³² and Kitaura and Morokuma,³³ among others.³⁴ Following the scheme of ref. 33, the SCF energy is divided into four components: (i) electrostatic; (ii) polarization; (iii) charge transfer; and (iv) exchange. When only one of the interacting molecules has a QM nature, only electrostatic and polarization contribute to the SCF energy.

An alternative to the SCF treatment of the interaction energy can be obtained from perturbation theory. In the particular case of a QM particle interacting with a single monopole placed at position r_1 , the interaction energy at the first-order perturbational level is given by the molecular electrostatic potential (MEP):

$$V(r_1) = \sum_A \frac{Z_A}{|r_1 - R_A|} - \sum_i^{\text{occ}} \sum_{\mu} \sum_{\nu} c_{\mu i} c_{\nu i} \langle \phi_{\mu} | \frac{1}{|r_1 - r|} | \phi_{\nu} \rangle \quad (6)$$

where V is the MEP, ϕ stands for the set of basis functions used for the QM molecule A , and $c_{\mu i}$ is the coefficient of atomic orbital μ in the molecular orbital i .

The MEP can be generalized to account for the electrostatic interaction with any classical charge distribution, taking advantage of the additivity of Coulombic interactions, which yields eq. (7), where the summation runs over all the charges (typically centered at atoms) in classical molecules B :

$$E_{\text{ele}} = \sum_B Q_B V(R_B) \quad (7)$$

The addition of a Lennard–Jones potential to Eq. (7) leads to the generalized molecular interaction potential (GMIP), which allows for fast and simple calculation of interaction energies between nonpolarizable QM and classical particles [see eq. (8)]. After careful parameterization of the Lennard–Jones parameters,²⁹ the GMIP provides a very accurate picture of intermolecular interactions at reduced computational cost:

$$\text{GMIP} = \sum_B Q_B V(R_B) + \sum_{A'B'} \left(\frac{C_{A'B'}}{|R_{B'} - R_{A'}|^{12}} - \frac{D_{A'B'}}{|R_{B'} - R_{A'}|^6} \right) \quad (8)$$

where the index B refers to charge centers in classical molecules, and A' and B' denote van der Waals centers in QM and classical molecules.

Polarization effects can be included at the second-order perturbation level [eq. (9)]. As shown by Francl,³¹ the interaction with a proton at the Hartree–Fock level can be determined approximately using eq. (10) (note a typo error in the equivalent expression of ref. 31). Eq. (10) has been implemented in the MEPHISTO³⁵ computer program, which allows the introduction of proton-induced polarization into QM particles:

$$E_{\text{pol}} = \langle \Psi^0 | \sum_B \frac{Q_B}{|r - R_B|} | \Psi^1 \rangle \quad (9)$$

$$E_{\text{pol}}(r_1) = \sum_j^{\text{vir}} \sum_i^{\text{occ}} \frac{1}{\varepsilon_i - \varepsilon_j} \times \left\{ \sum_{\mu} \sum_{\nu} c_{\mu i} c_{\nu j} \langle \phi_{\mu} | \frac{1}{|r_1 - r|} | \phi_{\nu} \rangle \right\}^2 \quad (10)$$

where ε stands for the energy of virtual (j) or occupied (i) molecular orbitals.

Eq. (10) can be easily extended to account for the interaction with a set of classical point charges, leading to eq. (11), which can be combined with the GMIP formalism to derive the *generalized molecular interaction potential with polarization correction* (GMIPp), as shown in eq. (12):

$$E_{\text{pol}} = \sum_j^{\text{vir}} \sum_i^{\text{occ}} \frac{1}{\varepsilon_i - \varepsilon_j} \times \left\{ \sum_{\mu} \sum_{\nu} c_{\mu i} c_{\nu j} \langle \phi_{\mu} | \sum_B \frac{Q_B}{|R_B - r|} | \phi_{\nu} \rangle \right\}^2 \quad (11)$$

$$\begin{aligned} \text{GMIPp} = & \sum_A \sum_B \frac{Q_B Z_A}{|R_B - R_A|} \\ & - \sum_B Q_B \sum_i^{\text{occ}} \sum_{\mu} \sum_{\nu} c_{\mu i} c_{\nu i} \langle \phi_{\mu} | \frac{1}{|R_B - r|} | \phi_{\nu} \rangle \\ & + \sum_{A'B'} \left(\frac{C_{A'B'}}{|R_{B'} - R_{A'}|^{12}} - \frac{D_{A'B'}}{|R_{B'} - R_{A'}|^6} \right) \\ & + \sum_j^{\text{vir}} \sum_i^{\text{occ}} \frac{1}{\varepsilon_i - \varepsilon_j} \\ & \times \left\{ \sum_{\mu} \sum_{\nu} c_{\mu i} c_{\nu j} \langle \phi_{\mu} | \sum_B \frac{Q_B}{|R_B - r|} | \phi_{\nu} \rangle \right\}^2 \end{aligned} \quad (12)$$

It is worth noting that eq. (12) allows the rigorous calculation of electrostatic interactions between QM and classical particles, the introduction of an empirical and parameterized term to represent dispersion–repulsion interactions, and the perturbative treatment of the polarization term. The quality of this latter component determined with different basis sets for a wide set of QM molecules and for various classical charge distributions will be examined in this study. Indeed, the reliability of the GMIPp as a general reactivity index, as well as its suitability as an effective Hamiltonian for mixed QM/MM calculations, will be discussed.

Computational Details

Molecular wave functions and SCF estimates of polarization energy were determined using Gaussian-94.³⁶ MEP, GMIP, and GMIPp calculations were done using our latest version of the MOPETE³⁷ computer program. All calculations were done at the HF level.

Monte Carlo simulations of solutes in aqueous solutions were performed using BOSS3.4³⁸ with the standard protocol at constant temperature (298 K) and pressure (1 atm). The TIP3P water model³⁹ was used. The solute was described using OPLS van der Waals parameters and charges when available, or 6-31G⁴⁰ electrostatic-potential-derived charges^{41–43} when OPLS charges were not available. The HF/6-31G* gas-phase-optimized geometries were used in all cases.

Calculations were performed in the SP2 computer of the Centre de Supercomputació de Catalunya (CESCA), as well as in workstations at our laboratory.

Results and Discussion

BASIS SET DEPENDENCE

To determine the dependence of the polarization energy on the basis set, the SCF interaction energy between 34 prototypical molecules (23 neutral, 6 anions, and 5 cations; see Fig. 1) and a single point charge was determined using the following basis: STO-3G, 3-21G, 6-31G*, 6-31 + G*, and 6-311 + + G**.^{40,44–48} The unit (positive or negative) charge was placed in suitable orientations at 3 Å from the molecule (see Fig. 1), where it was expected to largely polarize the QM molecule. Different locations were considered for some molecules.

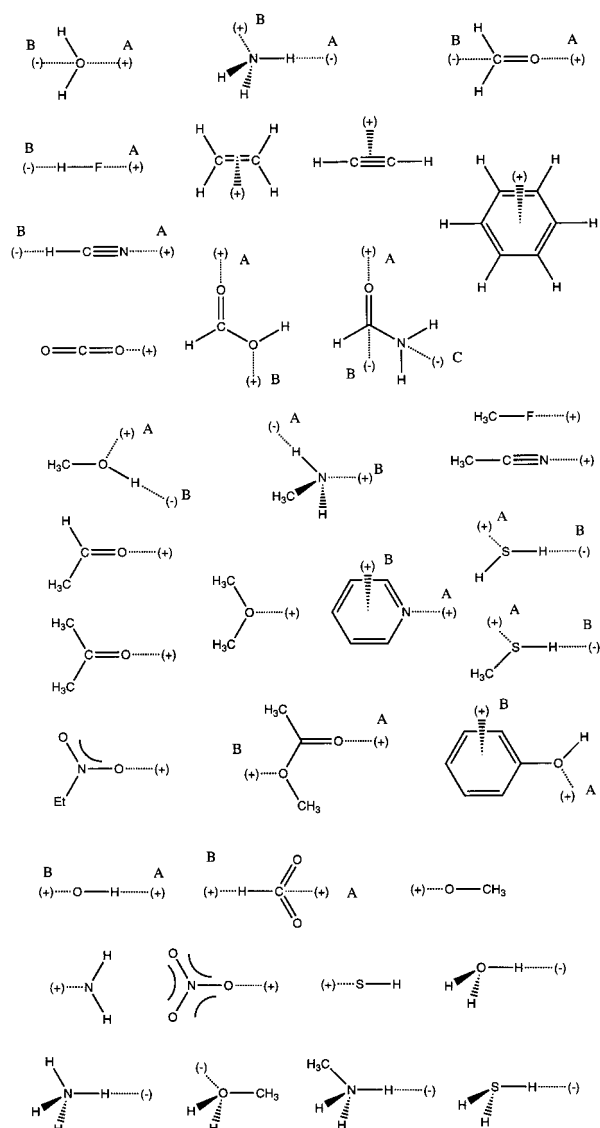


FIGURE 1. Schematic models for the interaction between a QM molecule (23 neutral, 6 anions, and 5 cations) and a single point charge placed at specific positions (denoted by capital letters).

Results in Table I show the large magnitude of the polarization contribution to the interaction energy, which in some cases is comparable to the electrostatic term. As expected, the polarization energy is low in cations, whereas anions are largely polarizable. Comparison of neutral and ionic species reveals that the magnitude of polarization effects increases from cation to neutral to anion for a given compound.

The polarization contribution can be very important for neutral molecules, because this term is in some cases similar or more negative than the

TABLE I.
Electrostatic (ele) and Polarization (pol) Interaction Energies (kcal / mol) between QM Molecules and a Point Charge Computed at the SCF Level for Different Basis Sets.^a

Molecule	Model	Energy	STO-3G	3-21G	6-31G*	6-31+G*	6-311++G**
H ₂ O	A	ele	-12.5	-17.6	-15.8	-16.6	-16.1
	A	pol	-0.5	-0.9	-1.3	-1.7	-1.7
	B	ele	-13.0	-17.8	-15.8	-16.5	-15.6
	B	pol	-0.5	-1.0	-1.4	-18.0	-2.0
NH ₃	A	ele	-3.9	-4.8	-4.7	-4.9	-4.7
	A	pol	-0.7	-1.2	-1.2	-1.3	-1.4
	B	ele	-15.9	-20.2	-19.1	-20.1	-18.9
	B	pol	-0.4	-1.0	-1.7	-3.2	-3.6
H ₂ CO	A	ele	-9.3	-14.9	-14.5	-15.5	-15.5
	A	pol	-1.4	-2.2	-2.6	-3.1	-3.1
	B	ele	-6.5	-13.1	-14.0	-14.9	-14.4
	B	pol	-1.6	-3.0	-3.2	-3.6	-3.8
HF	A	ele	-7.1	-12.0	-10.3	-10.6	-10.7
	A	pol	-0.6	-0.7	-0.9	-1.1	-1.1
	B	ele	-7.8	-13.0	-12.5	-13.2	-12.9
	B	pol	-0.6	-0.6	-0.6	-0.6	-0.8
HCN		ele	-13.6	-16.0	-15.7	-15.9	-15.9
		pol	-1.7	-2.9	-3.4	-4.1	-4.2
C ₂ H ₄		ele	-3.1	-7.2	-7.0	-8.0	-7.9
		pol	-0.8	-2.0	-2.9	-6.8	-8.1
C ₂ H ₂		ele	-4.2	-6.8	-7.0	-7.3	-7.4
		pol	-0.7	-1.7	-2.5	-4.0	-4.5
C ₆ H ₆		ele	-4.8	-12.6	-11.0	-11.2	-11.4
		pol	-1.6	-3.5	-5.0	-8.8	-9.2
CO ₂		ele	-5.6	-7.3	-5.4	-5.5	-5.5
		pol	-1.5	-2.1	-2.4	-2.8	-2.9
HCOOH	A	ele	-9.0	-13.6	-13.4	-14.0	-14.1
	A	pol	-1.4	-2.2	-2.5	-3.1	-3.0
	B	ele	-4.4	-5.2	-3.0	-3.1	-3.0
	B	pol	-1.0	-1.7	-2.1	-2.9	-2.7
HCONH ₂	A	ele	-14.2	-20.5	-20.0	-21.0	-20.9
	A	pol	-1.7	-2.5	-2.9	-3.6	-3.7
	B	ele	-21.4	-30.8	-31.7	-33.4	-32.7
	B	pol	-6.7	-8.8	-9.5	-10.2	-10.7
	C	ele	-10.3	-14.5	-15.4	-16.2	-15.9
	C	pol	-1.9	-2.9	-3.4	-3.7	-3.9
CH ₃ OH	A	ele	-11.4	-16.2	-14.5	-15.6	-15.3
	A	pol	-0.8	-1.6	-2.1	-2.7	-2.7
	B	ele	-6.0	-8.1	-7.9	-8.3	-8.0
	B	pol	-0.9	-1.3	-1.4	-1.4	-1.5
CH ₃ NH ₂	A	ele	-7.7	-9.5	-9.7	-10.2	-9.8
	A	pol	-3.6	-5.3	-5.6	-5.8	-6.1
	B	ele	-14.5	-18.4	-17.5	-18.9	-18.0
	B	pol	-0.8	-1.8	-2.5	-4.8	-5.3
CH ₃ F		ele	-6.2	-11.9	-10.3	-11.4	-11.7
		pol	-1.0	-1.4	-1.7	-2.0	-2.0
CH ₃ CN		ele	-16.4	-19.4	-19.4	-19.8	-19.9
		pol	-1.8	-3.2	-3.7	-4.7	-4.8
CH ₃ CHO		ele	-10.5	-16.3	-16.0	-17.2	-17.2
		pol	-1.6	-2.6	-3.0	-3.5	-3.6
CO(CH ₃) ₂		ele	-11.3	-17.3	-17.1	-18.3	-18.4
		pol	-1.8	-2.9	-3.3	-4.0	-4.1

TABLE I.
(Continued).

Molecule	Model	Energy	STO-3G	3-21G	6-31G*	6-31+G*	6-311++G**
O(CH ₃) ₂		ele	-10.0	-14.4	-12.5	-13.4	-13.0
		pol	-1.1	-2.1	-2.6	-3.5	-3.6
Pyridine	A	ele	-14.7	-17.0	-16.4	-17.1	-17.0
	A	pol	-2.3	-3.9	-4.8	-7.1	-7.4
	B	ele	-4.3	-9.0	-8.8	-9.0	-8.9
	B	pol	-2.1	-3.8	-4.9	-7.9	-8.3
H ₂ S	A	ele	-7.3	-12.2	-8.4	-8.4	-7.9
	A	pol	-1.5	-3.7	-4.2	-4.5	-4.6
	B	ele	-3.4	-6.4	-5.7	-5.8	-5.8
	B	pol	-1.2	-2.1	-2.0	-2.1	-2.1
CH ₃ SH	A	ele	-7.4	-13.7	-10.9	-11.1	-10.7
	A	pol	-2.0	-5.0	-5.7	-6.4	-6.7
	B	ele	-2.8	-4.6	-4.2	-4.4	-4.4
	B	pol	-1.4	-2.5	-2.4	-2.5	-2.5
CH ₃ CH ₂ NO ₂		ele	-10.4	-13.1	-12.1	-12.4	-12.7
		pol	-2.1	-3.1	-3.4	-3.9	-4.0
CH ₃ COOCH ₃	A	ele	-10.5	-15.5	-15.4	-16.0	-16.0
	A	pol	-1.8	-2.8	-3.2	-4.0	-4.1
	B	ele	-3.3	-4.2	-2.1	-2.2	-2.4
	B	pol	-1.9	-3.3	-3.9	-5.2	-5.4
C ₆ H ₅ OH	A	ele	-8.5	-11.1	-10.0	-10.3	-10.1
	A	pol	-2.5	-3.7	-4.3	-5.3	-5.4
	B	ele	-4.3	-11.5	-10.3	-10.3	-10.5
	B	pol	-1.7	-3.7	-5.1	-9.2	-9.7
OH ⁻	A	ele	-109.1	-114.0	-114.2	-115.5	-114.9
	A	pol	-0.8	-1.6	-1.9	-2.6	-2.9
	B	ele	-85.1	-79.4	-77.9	-74.1	-75.1
	B	pol	-0.7	-1.4	-1.7	-1.7	-2.4
HCOO ⁻	A	ele	-102.1	-104.5	-105.4	-107.0	-107.0
	A	pol	-0.6	-1.3	-2.2	-4.4	-4.6
	B	ele	-71.5	-68.5	-67.9	-65.7	-65.8
	B	pol	-1.0	-2.4	-2.5	-3.3	-4.0
CH ₃ O ⁻		ele	-99.5	-102.7	-103.5	-106.9	-106.6
		pol	-1.2	-2.3	-2.7	-3.9	-4.2
NH ₂ ⁻		ele	-109.9	-107.9	-106.5	-102.0	-102.3
		pol	-1.9	-4.0	-4.5	-4.8	-5.8
NO ₃ ⁻		ele	-84.2	-84.6	-83.8	-83.6	-83.8
		pol	-1.6	-2.5	-2.9	-3.7	-3.7
SH ⁻		ele	-106.5	-108.2	-105.7	-103.7	-103.5
		pol	-1.7	-4.8	-5.5	-6.3	-7.0
H ₃ O ⁺		ele	-88.6	-89.3	-89.0	-89.1	-89.0
		pol	-0.6	-0.7	-0.7	-0.7	-0.8
NH ₄ ⁺		ele	-84.3	-84.4	-84.4	-84.4	-84.4
		pol	-0.7	+0.9	-1.0	-1.0	-1.1
CH ₃ OH ₂ ⁺		ele	-104.3	-103.2	-103.8	-103.9	-103.9
		pol	-2.0	-2.2	-2.6	-2.8	-2.9
CH ₃ NH ₃ ⁺		ele	-77.5	-77.1	-77.2	-77.2	-77.1
		pol	-0.8	-1.0	-1.0	-1.1	-1.2
H ₃ S ⁺		ele	-82.3	-83.0	-82.0	-82.0	-82.1
		pol	-1.2	-1.8	-1.6	-1.7	-1.6

^aPoint charges (positive or negative depending on the model; see Fig. 1) are placed at 3 Å from the molecule.

electrostatic energy (see Table I). The polarizability of molecules is enlarged by the addition of methyl groups and by the presence of π -aromatic systems.⁴⁹ In this latter case, the polarization energy is particularly large. This warns against the use of classical nonpolarizable force fields to study these systems. It is worth noting the results reported for pyridine and phenol, where the polarization energy for the charge- π interaction is 1–4 kcal/mol more negative than for the charge- σ interaction when the 6-311++G** basis is used. This finding demonstrates the complexity and anisotropy of polarization effects, and illustrates the difficulty in properly reproducing the polarization from atom-based parameters.

Comparison of the polarization energy determined with the different basis is shown in Figure 2. In all cases, the values obtained with the 6-311++G** basis were used as reference. The results clearly show the inadequacy of minimum and small split-valence basis, because the polarization energy is clearly underestimated in a nonsystematic way. Better agreement is found for the medium-sized 6-31G* basis, but only inclusion of diffuse functions leads to qualitatively correct results. Inspection of the plot for the 6-31G* results reveals that most of the discrepancy stems from a few outliers, which correspond to charge- π interactions. The statistical results improve reasonably upon exclusion of these interactions from the com-

parison study (see Fig. 3). Thus, the root-mean-square (RMS) error for the 6-31G* basis decreases to only 0.9 kcal/mol. It must be emphasized that the errors are expected to be much smaller when interactions with classical neutral molecules are considered.

In summary, the analysis suggests that the 6-31+G* basis is the minimum level necessary to recover properly the polarization of molecules. However, if interactions with π -charge distributions are excluded, the 6-31G* basis is accurate enough to reproduce the polarization effects. The accuracy of the results determined with this basis is even larger when only neutral and cation species are considered (data not shown). Finally, minimum or small split-valence basis sets are strongly discouraged, because the polarization energy values show poor correlations. The 6-31G* and occasionally 6-31+G* basis sets are used throughout the rest of the article.

PERTURBATIONAL TREATMENT OF SINGLE POINT CHARGE POLARIZATION

The ability of the perturbational treatment [PT; eq. (11)] to capture the polarization energy in interactions with single point charges is examined here. The SCF and PT results for the entire set of interac-

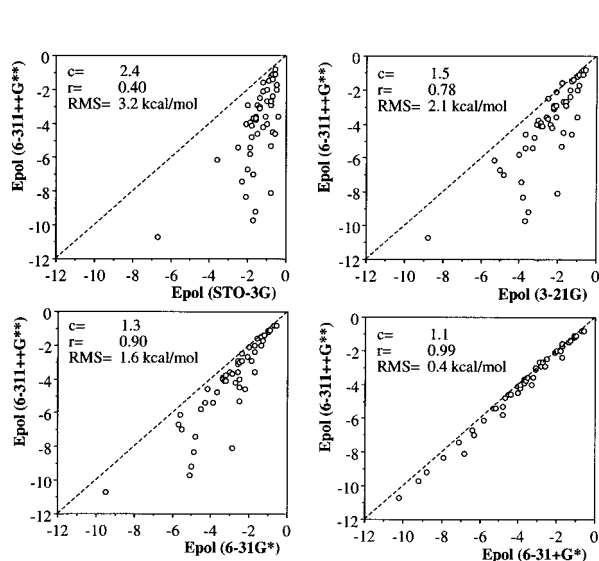


FIGURE 2. Comparison of the SCF polarization energy (kcal/mol) determined with the basis STO-3G, 3-21G, 6-31G*, and 6-31+G* with regard to the SCF values computed using the 6-311++G** basis. All the models in Figure 1 are considered. The dashed line shows the ideal regression ($c = 1.0$).

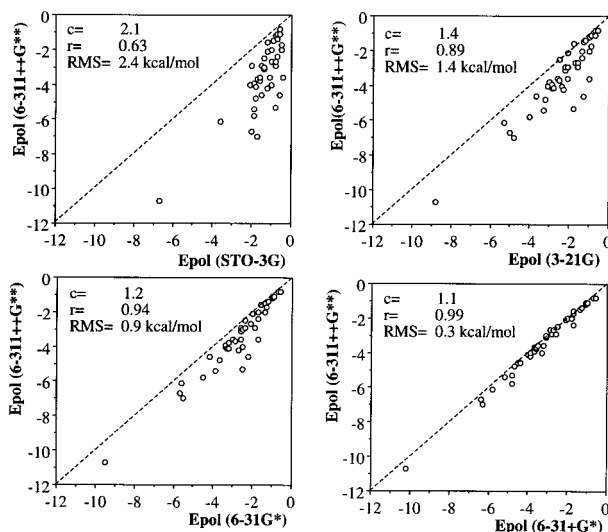


FIGURE 3. Comparison of the SCF polarization energy (kcal/mol) determined with the basis STO-3G, 3-21G, 6-31G*, and 6-31+G* with regard to the SCF values computed using the 6-311++G** basis. Charge- π interactions models in Figure 1 were excluded from the comparison. The dashed line shows the ideal regression ($c = 1.0$).

tions (see Fig. 1) are given in Table II, and Figure 4 shows the statistical results of the comparison.

Inspection of Table II and Figure 4 clearly shows that SCF and PT results are nearly identical, the RMS error being only 0.4 kcal/mol. In most cases, the difference between SCF and PT values is less than 0.2 kcal/mol, and only a few interactions show differences larger than 0.5 kcal/mol. Interactions with π -charge distributions are well reproduced at the PT level, the largest differences being found for a few anions and for sulfur-containing compounds. In any case, it is worth noting that errors due to the PT treatment are much smaller than those arising from the use of a medium-sized basis set. Therefore, these results strongly support the use of eq. (11) as a fast, inexpensive computational technique to compute polarization in mixed QM/MM methods.

SINGLE POINT CHARGE POLARIZATION EFFECTS IN THE SPACE

In the previous subsection we analyzed the polarization induced by a classical point charge placed in specific positions relative to the QM molecule. A point of interest is to analyze the dependence of the polarization on the distance between the QM and classical particles, and to examine the reliability of the PT approach for different configurations of the QM classical system. Following our systematic approach, we first considered the simplest case: a point charge (positive or negative) interacting at different distances for selected molecules. Plots of the SCF and PT polarization energy are shown in Figure 5.

As expected, polarization is maximum in regions near the nuclei and decays at large distances, being very small (< 0.3 kcal/mol) for distances greater than 6 Å. At short distances (2 Å) polarization can in some cases be larger than 15 kcal/mol (in absolute values), which demonstrates the importance of this term in inner regions. In the range of 3–5 Å, where weak chemical interactions occur, the polarization energy is far from being negligible. For instance, in this range of distances, polarization varies from -1.3 to -0.1 kcal/mol even for low polarizable molecules like H_2O and H_3O^+ . As expected, the molecules containing π -electron systems are more polarizable in the whole range of distance, the polarization being almost 1 kcal/mol at a distance of 5 Å. Finally, polarization is also high in inner regions of H_2S , as a consequence of the larger size of the sulfur atom. However, polarization decays rapidly with the distance, and for

TABLE II.
SCF and PT Estimates of Polarization Interaction Energy (kcal / mol) between QM Molecules and a Point Charge Computed Using the 6-31G* Basis Set.^a

Molecule	Model	SCF	PT
H_2O	A	-1.3	-1.1
	B	-1.4	-1.2
NH_3	A	-1.2	-1.0
	B	-1.7	-1.3
H_2CO	A	-2.6	-2.4
	B	-3.2	-3.1
HF	A	-0.9	-0.9
	B	-0.6	-0.5
HCN		-3.4	-3.7
C_2H_4		-2.9	-2.6
C_2H_2		-2.5	-2.1
C_6H_6		-5.0	-5.3
CO_2		-2.4	-2.5
HCOOH	A	-2.5	-2.5
	B	-2.1	-2.0
HCONH_2	A	-2.9	-2.7
	B	-9.5	-8.7
	C	-3.4	-2.8
CH_3OH	A	-2.1	-1.8
	B	-1.4	-1.1
CH_3NH_2	A	-5.6	-4.7
	B	-2.5	-2.1
CH_3F		-1.7	-1.4
CH_3CN		-3.7	-3.6
CH_3CHO		-3.0	-2.7
$\text{CO}(\text{CH}_3)_2$		-3.3	-3.0
$\text{O}(\text{CH}_3)_2$		-2.6	-2.4
Pyridine	A	-4.8	-4.9
	B	-4.9	-5.2
H_2S	A	-4.2	-3.4
	B	-2.0	-1.7
CH_3SH	A	-5.7	-4.6
	B	-2.4	-2.0
$\text{CH}_3\text{CH}_2\text{NO}_2$		-3.4	-3.2
$\text{CH}_3\text{COOCH}_3$	A	-3.2	-3.0
	B	-3.9	-3.9
$\text{C}_6\text{H}_5\text{OH}$	A	-4.3	-4.1
	B	-5.1	-5.6
OH^-	A	-1.9	-1.6
	B	-1.7	-1.1
HCOO^-	A	-2.2	-2.0
	B	-2.5	-1.8
CH_3O^-		-2.7	-2.4
NH_2^-		-4.5	-3.2
NO_3^-		-2.9	-2.7
SH^-		-5.5	-4.6
H_3O^+		-0.7	-0.6
NH_4^+		-1.0	-0.9
CH_3OH_2^+		-2.6	-2.2
CH_3NH_3^+		-1.0	-0.9
H_3S^+		-1.6	-1.4

^aPoint charges (positive or negative depending on the model; see Fig. 1) are placed at 3 Å from the molecule.

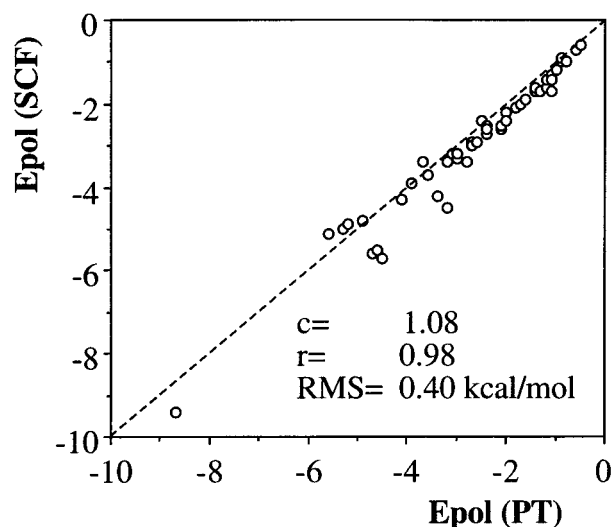


FIGURE 4. Comparison of the SCF and PT polarization energies (kcal/mol) determined for the models in Figure 1 with the 6-31G* basis. The dashed line shows the ideal regression ($c = 1.0$).

values of 5 Å the polarization energy is similar to that found for H₂O.

The dependence of the polarization with the third power of the distance is shown in Figure 6, which shows the fitting of the polarization energy to a simple r^{-3} equation ($r > 0.99$ in all cases). The weaker or stronger dependence with the distance reflects the differences in polarizability of the

compounds. The behavior of H₂S is anomalous, because the regression line crosses the lines obtained for the other molecules. The dependence of the polarization energy for this molecule cannot be described using a linear equation in the full range of distances. It seems that the larger size of the sulfur atom is responsible of this special behavior.

Figure 5 shows the reliability of the PT treatment in the whole range of distances. The largest discrepancies appear in the inner regions, where

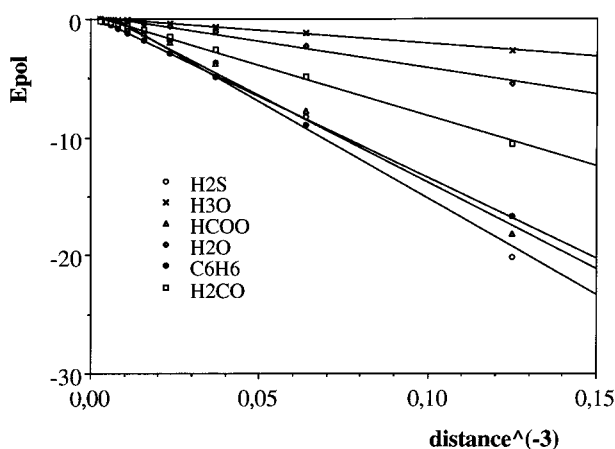


FIGURE 6. Dependence of the polarization energy (kcal/mol) on the inverse of the third power of the distance (Å⁻³) between the interacting QM molecule and the monopole.

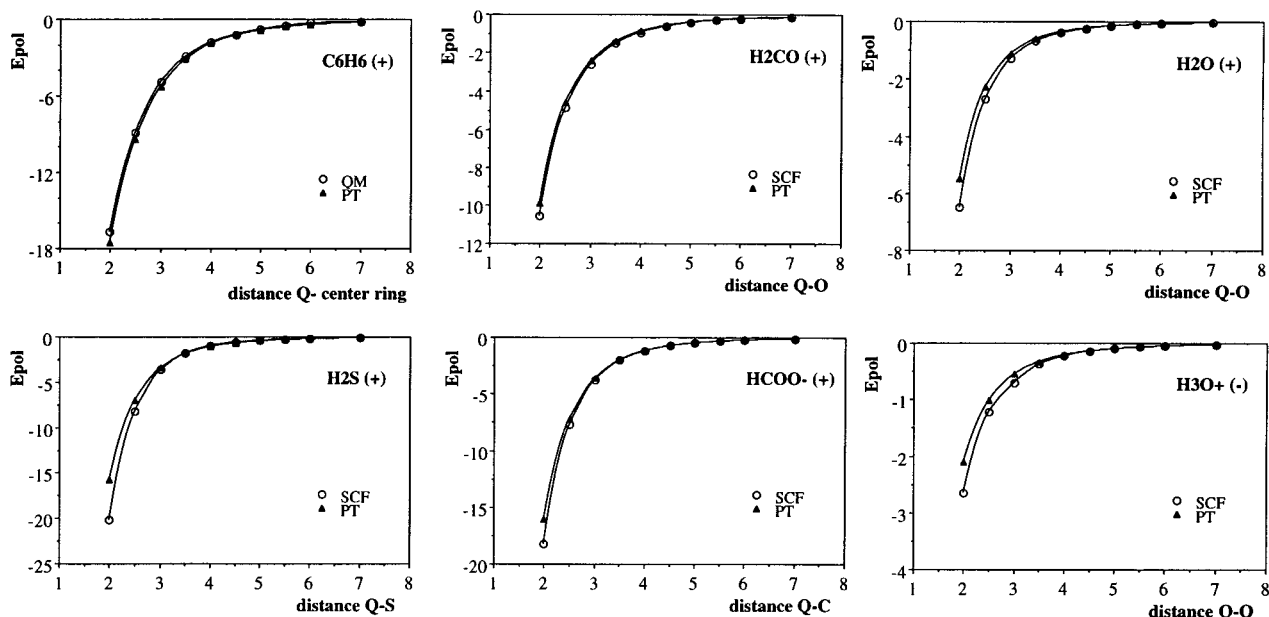


FIGURE 5. SCF and PT polarization energies (kcal/mol) for the interaction between a QM molecule and a single point charge placed at different distances (Å) from the molecule. The nature (positive or negative) of the interacting monopole is given in parentheses.

the polarization energy is underestimated by 10–20%. However, in the region of chemical interest, and obviously at larger distances, both SCF and PT results are nearly identical. Again, lower errors are expected for the polarization induced by classical molecules.

MULTIPOLE-INDUCED POLARIZATION EFFECTS

One of the main concerns regarding the use of eq. (11) is its reliability in treating the nonadditivity of polarization effects. This is necessary for an accurate representation of the polarization induced by a set of classical charges pertaining to one or more molecules. To examine this point, the SCF and PT polarization energies for the interaction of a set of unit point charges placed in key positions around a given QM molecule were determined (see Fig. 7 and Table III). The results demonstrate the preciseness of the PT values even in cases of large nonadditivity. The PT approach underestimates the polarization energy slightly and consistently by a factor of 9% for the molecules considered here (see Fig. 8). This deviation is much smaller than that due to the use of a medium-sized basis like the 6-31G*, and might lead to relevant discrepancies only for very unrealistic systems, like models III and IV of formamide, where the polarization reaches values of around -30

kcal/mol. Thus, when these two models are excluded from the analysis, the results deviate only 3% and the RMS error decreases to 0.7 kcal/mol.

The preceding results show that the reliability of the PT treatment of interactions with large non-additive polarization effects. However, the models chosen in Figure 7 do not reflect interactions in real chemical systems. To complete this study, the polarization energy of QM molecules interacting with more realistic chemical species represented

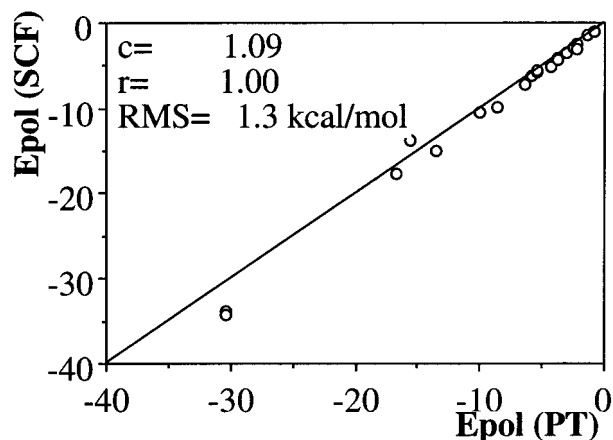


FIGURE 8. Comparison of the SCF and PT polarization energies (kcal/mol) determined for the models in Figure 7 with the 6-31G* basis. The plot also shows the ideal regression line ($c = 1.0$).

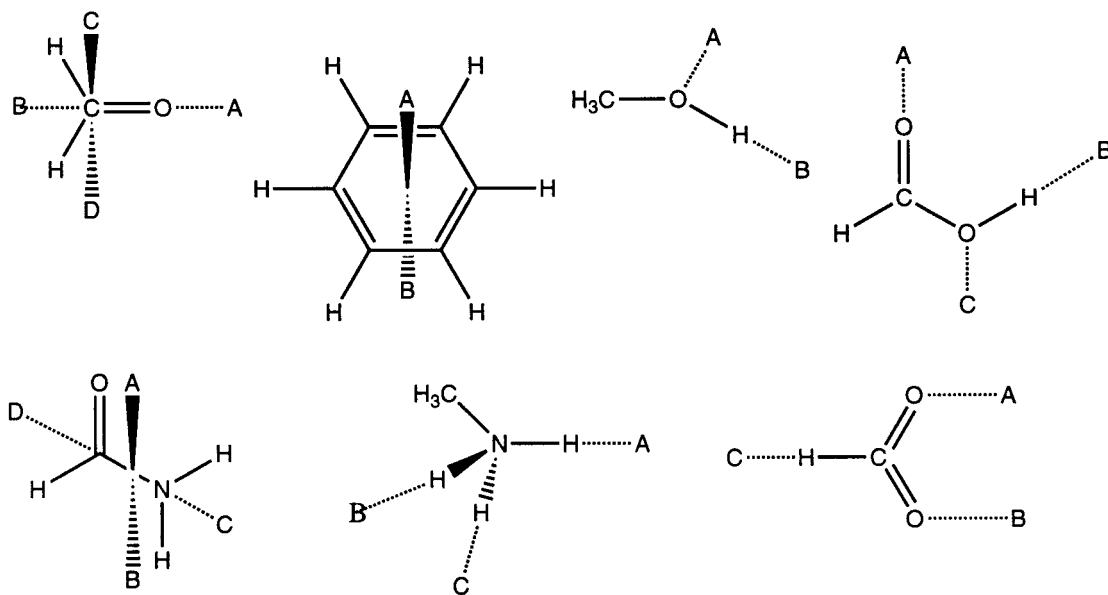


FIGURE 7. Schematic models for the interaction between a QM molecule and a multipole. The position of the set of point charges is denoted by capital letters (see also Table III).

by a set of classical monopoles was determined (see Fig. 9 and Table IV). As expected, the magnitude of the polarization energy is sensibly lower than the values found for interaction with the unit charge. However, it is clear that the polarization energy is not negligible in most cases. Results in Table IV also show the good performance of the PT approach, which yields values typically slightly less negative than the SCF values.

POLARIZATION EFFECTS IN CONDENSED PHASES

One of our goals is to design an effective Hamiltonian that combines accuracy and efficiency in QM/MM calculations. A challenging test of the validity of the PT approach is to evaluate the solvent-induced polarization of solutes. In particular, we compared the SCF and PT estimates of the polarization energy for systems consisting of one QM solute (HCOOH, H₂O, H₂CO, HCONH₂, H₃O⁺, and HCOO⁻) and 215 TIP3P water molecules. Owing to computational limitations,

Monte Carlo samplings were performed using classical force fields (see “Theoretical Framework” section). The polarization energy was determined for five or six configurations of the system, which were selected every 2 · 10⁵ configurations after 0.5 million configurations of equilibration. Estimates of the polarization energy were determined by averaging over the selected configurations. It is worth noting that the use of classical force fields to sample the configurational space is not expected to introduce any bias in the comparison analysis.

The results derived for the interaction with a single point charge indicated that polarization effects were negligible beyond a certain distance from the molecule (see previous sections). Therefore, the determination of the “polarization-cutoff” distance is very important, because its use can reduce drastically the cost of the QM/MM calculation. Accordingly, the comparison between SCF and PT polarization energies was performed varying the “cutoff” distance from 3.5 to 9 Å (this latter value was the “cutoff” used in Monte Carlo simulations). This allows us to examine not only

TABLE III. SCF and PT Estimates of Polarization Interaction Energy (kcal / mol) between QM Molecules and a Set of Point Unit Charges Computed Using the 6-31G* Basis.^a

Molecule	Model	A	B	C	D	SCF	PT
H ₂ CO	I	(+)	(-)			-10.5	-10.1
	II	(+)	(+)			-1.0	-0.9
	III	(+)		(-)		-5.6	-5.4
	IV	(+)		(-)	(-)	-6.2	-5.9
	V	(+)	(+)	(+)	(+)	-1.4	-1.4
	VI	(+)	(-)	(+)	(+)	-9.8	-8.6
	VII	(+)	(-)	(+)	(-)	-17.7	-16.7
C ₆ H ₆	I	(+)	(+)			-5.7	-5.5
	II	(+)	(-)			-13.7	-15.6
CH ₃ OH	I	(+)	(-)			-4.1	-3.7
	II	(-)	(+)			-4.3	-3.7
HCOOH	I	(-)	(+)			-3.5	-3.0
	II	(+)	(-)			-3.4	-3.0
	III	(+)	(-)	(+)		-2.9	-2.5
HCONH ₂	I	(+)	(+)			-2.4	-2.2
	II	(+)	(+)	(+)	(+)	-5.1	-4.3
	III ^b	(+)	(-)	(+)	(-)	-34.3	-30.4
	IV ^b	(+)	(-)	(-)	(+)	-33.8	-30.4
HCOO ⁻	I	(+)	(+)			-7.1	-6.4
	II	(+)	(+)	(-)		-15.0	-13.5
CH ₃ NH ₃ ⁺		(-)	(-)	(-)		-3.0	-2.2

^aPoint charges (positive or negative depending on the model; see Fig. 7) are placed at 3 Å from the molecule.
^bIn these models, charges C and D were placed at 2.5 Å from the molecule.

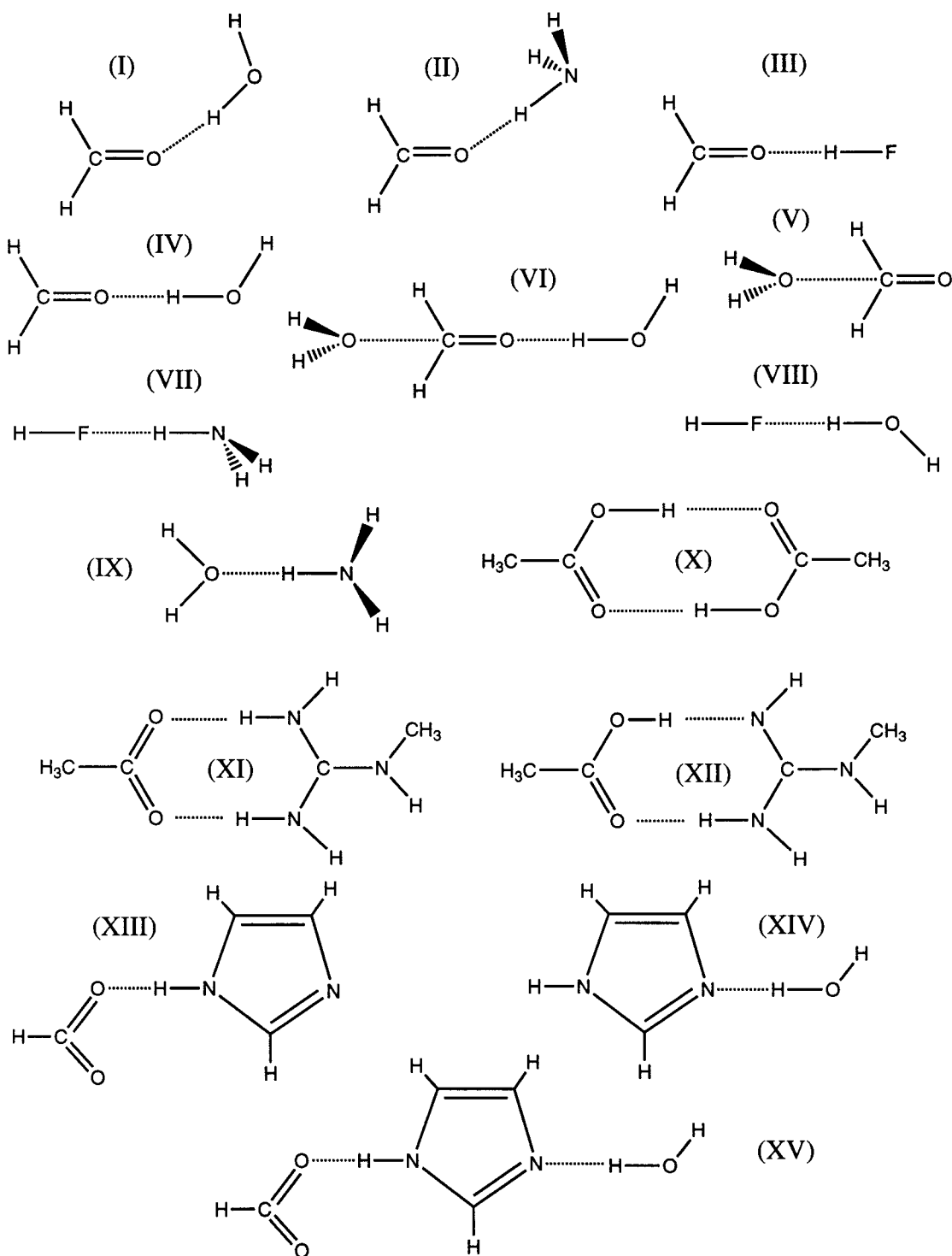


FIGURE 9. Schematic models for the interaction between a QM molecule and a classical molecule treated by a set of point charges (see also Table IV).

TABLE IV.
SCF and PT Estimates of Polarization Energy (kcal / mol) for Interactions between QM Molecules with
Classical Molecules Treated as a Set of Multiples Computed Using the 6-31G* Basis (see Fig. 9).^a

System	QM	Classical	SCF	PT
I	H ₂ CO	H ₂ O	-0.2	-0.2
	H ₂ O	H ₂ CO	-0.2	-0.2
II	H ₂ CO	NH ₃	-0.1	-0.1
	NH ₃	H ₂ CO	-0.1	-0.1
III	H ₂ CO	HF	-0.7	-0.6
	HF	H ₂ CO	-0.2	-0.2
IV	H ₂ CO	H ₂ O	-0.3	-0.3
V	H ₂ CO	H ₂ O	-0.3	-0.3
VI	H ₂ CO	H ₂ O	-1.0	-0.9
VII	HF	NH ₃	-0.1	-0.1
	NH ₃	HF	-0.2	-0.2
VIII	HF	H ₂ O	-0.2	-0.1
	H ₂ O	HF	-0.1	-0.1
IX	H ₂ O	NH ₃	-0.1	-0.1
	NH ₃	H ₂ O	-0.4	-0.3
X	CH ₃ COOH	CH ₃ COOH	-1.9	-1.6
XI	Acetate	Guanidinium	-5.3	-4.6
	Guanidinium	Acetate	-9.0	-7.9
XII	Acetic acid	Guanidine	-1.8	-1.4
	Guanidine	Acetic acid	-2.0	-1.5
XIII	Imidazole	Formate	-5.7	-5.4
	Formate	Imidazole	-0.6	-0.5
XIV	Imidazole	H ₂ O	-0.4	-0.4
	H ₂ O	Imidazole	-0.4	-0.3
XV	Imidazole	Formate, H ₂ O	-7.5	-7.2

^aThe intermolecular distance is ~ 3 Å.

the reliability of the PT approach for increasing number of solvent molecules, but also to determine the value of the “polarization cutoff.”

The polarization of neutral molecules can be up to -2 kcal/mol (see Fig. 10) in polar π -rich systems. It is clear that such a large contribution to the total free energy of solvation cannot be neglected. Polarization effects are greater for anions (more than -5 kcal/mol for HCOO⁻), and quite small for cations (around 0.4 kcal/mol for H₃O⁺).

In general, polarization of cations and neutral molecules reaches convergence at distances beyond 6 Å. The only exception is formamide, for which convergence is not fully guaranteed in the range of distances considered. With respect to anions, polarization is not converged at such a distance from the molecule, and more external shells of water should still be considered.

Finally, it is also worth noting that SCF and PT estimates are very similar. Thus, the average difference is around 0.15 kcal/mol. The largest differ-

ence is found for HCOO⁻ (0.4 kcal/mol), but again this deviation is lower than the error originated from the inclusion of a small number of water molecules, or to the use of basis sets with limited flexibility. In summary, the present results support the validity of eq. (11) for treatment of polarization effects in condensed phases, and provide confidence in the use of the GMIPp as an effective Hamiltonian for mixed QM/MM calculations.

COMPUTATIONAL REQUIREMENTS

The use of eq. (11) implies the calculation of monoelectron integrals, which are in fact necessary for the determination of the electrostatic interaction (MEP). Therefore, inclusion of polarization effects with the GMIPp can be performed with a moderate increase in the computational cost. In fact, the GMIPp allows a significant saving in CPU time compared to the calculation of the polarization contribution using SCF-based procedures,

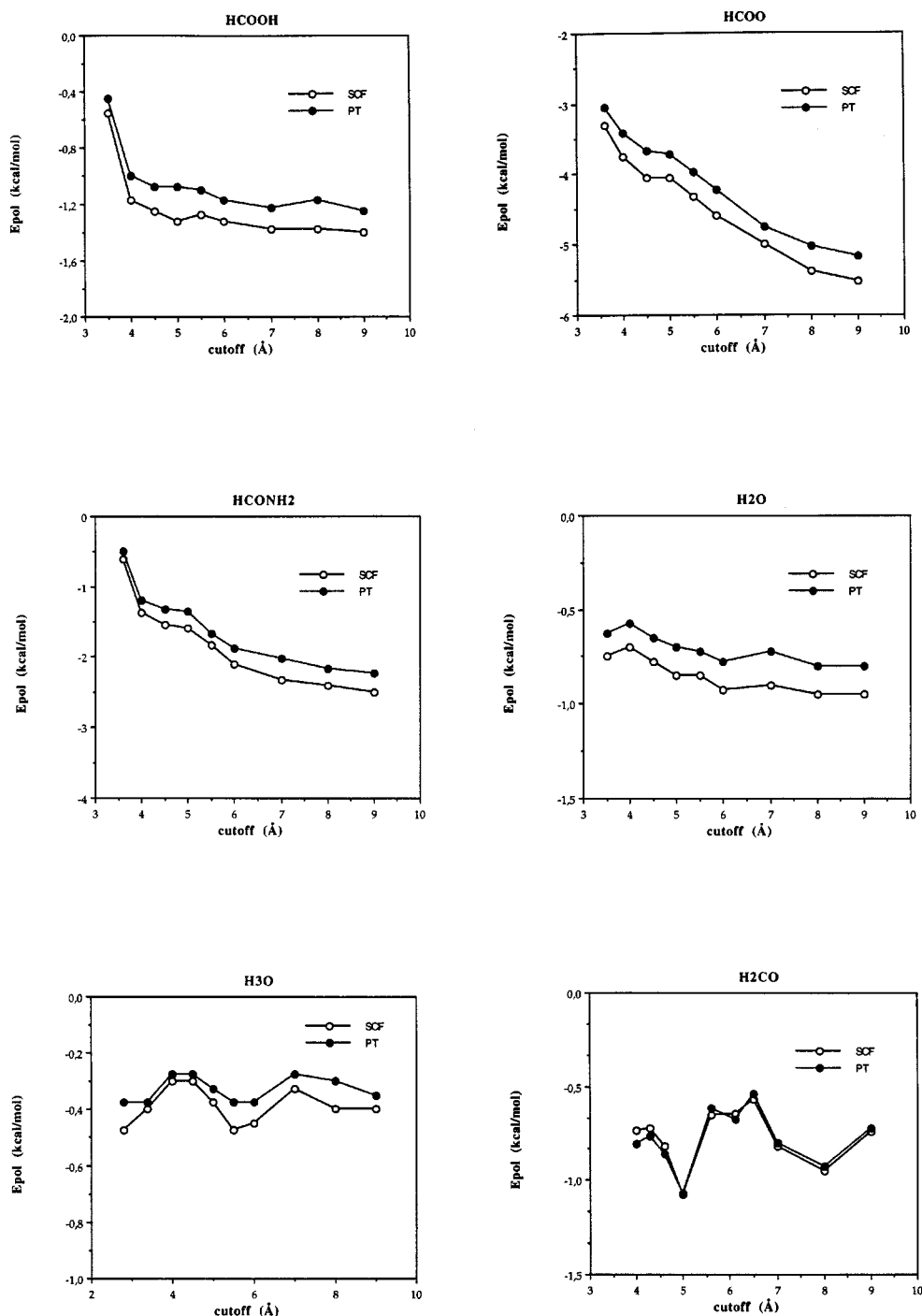


FIGURE 10. Representation of the SCF and PT polarization energy (kcal/mol) for selected QM solutes in aqueous solution computed using the 6-31G* (neutral) and 6-31+G* (charged) basis sets. The polarization was determined excluding those water molecules of the simulation system placed beyond a certain cutoff distance (Å) from the QM molecule.

since in these methods bielectronic integrals have to be recomputed, or retrieved from disk, in order to perform several self-consistent field cycles.

The GMIPp calculations reported here have been performed with the MOPETE program. The code is not optimized and its performance could be sensibly improved upon implementation of very efficient computational algorithms for calculation of monoelectron integrals.^{50,51} However, the CPU time needed for a GMIPp calculation compares very well with that required to compute the polarization in test systems using the Gaussian code. Thus, the ratio of CPU times required for determining the polarization in the interaction between phenol and the unit charge with MOPETE and Gaussian is 1:29 (using the default options for calculation of integrals and SCF convergence). Note that even when the molecular wave function is used as a guess, the ratio is 1:24. The CPU time of the PT calculation can still be reduced using several strategies. For instance, if the frozen-core approximation is used in the PT approach, the ratio with regard to the standard calculation in MOPETE is 0.7 with negligible changes in the magnitude of the PT polarization energy. Further improvements in the performance of the PT calculations can still be envisaged by using an integral cutoff.⁵¹

Conclusions

The present results suggest that the GMIPp appears to be a very promising tool for the study of interactions between molecules in systems where partitioning between QM and classical portions can be performed. The GMIPp is able to capture polarization effects accurately. In fact, our results indicate that the errors due to the use of the PT approach are sensibly smaller than those arising from the use of medium-sized basis sets. Indeed, calculation of the GMIPp is performed at a low computational cost, which makes this approach very promising for reactivity studies, as well as for QM/MM calculations.

Acknowledgments

We thank Prof. W. L. Jorgensen for a copy of his BOSS3.4 computer program. We are also indebted to Dr. J. J. Pérez and Dr. J. Gao for helpful discussions.

References

1. J. Gao, *Reviews in Computational Chemistry*, Vol. 7, K. B. Lipkowitz and D. B. Boyd, Eds., VCH, New York, 1996, p. 119.
2. J. Aqvist and A. Warshel, *Chem. Rev.*, **93**, 2523 (1993).
3. J. Gao, F. J. Luque, and M. Orozco, *J. Chem. Phys.*, **98**, 2975 (1993).
4. F. J. Luque, C. Alhambra, and M. Orozco, *J. Phys. Chem.*, **99**, 11344 (1995).
5. M. Orozco, F. J. Luque, D. Habibollahzadeh, and J. Gao, *J. Chem. Phys.*, **102**, 6145 (1995).
6. J. Mestres, M. Solà, R. Carbó, F. J. Luque, and M. Orozco, *J. Phys. Chem.*, **100**, 606 (1996).
7. M. Orozco, C. Colomines, and F. Luque, *J. Chem. Phys.*, **209**, 19 (1996).
8. C. J. Bottcher, *Theory of Electric Polarization*, Elsevier, Amsterdam, 1973.
9. A. Warshel and M. Levitt, *J. Mol. Biol.*, **103**, 227 (1976).
10. M. A. Thompson and G. K. Schenter, *J. Phys. Chem.*, **99**, 6374 (1995).
11. J. Gao and C. Alhambra, *J. Am. Chem. Soc.*, **119**, 2962 (1997).
12. J. Gao, *J. Comput. Chem.*, **18**, 1062 (1997).
13. J. Gao and K. Byun, *Theor. Chem. Acc.*, **96**, 151 (1997).
14. M. J. Field, P. A. Bash, and M. Karplus, *J. Comput. Chem.*, **11**, 700 (1990).
15. V. Luzhkov and A. Warshel, *J. Comput. Chem.*, **12**, 199 (1992).
16. O. Tapia, F. Colonna, and J. G. Angyan, *J. Chim. Phys.*, **87**, 875 (1990).
17. J. Gao and X. Xia, *Science*, **258**, 631 (1992).
18. R. V. Stanton, D. S. Hartsough, and K. M. Merz Jr., *J. Comput. Chem.*, **16**, 113 (1995).
19. R. V. Stanton, L. R. Little, and K. M. Merz, Jr., *J. Phys. Chem.*, **99**, 483 (1995).
20. D. Bakowies and W. Thiel, *J. Phys. Chem.*, **100**, 10580 (1996).
21. R. P. Muller and A. Warshel, *J. Phys. Chem.*, **99**, 17516 (1995).
22. T. Wesolowski and A. Warshel, *J. Phys. Chem.*, **98**, 5183 (1994).
23. V. Thery, D. Rinaldi, J. L. Rivail, B. Maigret, and G. G. Ferenczy, *J. Comput. Chem.*, **15**, 269 (1994).
24. F. Maseras and M. Morokuma, *J. Comput. Chem.*, **16**, 1170 (1995).
25. H. Liu, F. Müller-Plathe, and W. F. van Gunsteren, *J. Chem. Phys.*, **102**, 1722 (1995).
26. A. Broo, G. Pearl and M. C. Zerner, *J. Phys. Chem.*, **101**, 2478 (1997).
27. I. Tuñón, M. T. C. Martins-Costa, C. Millot, M. F. Ruiz-López, and J. L. Rivail, *J. Comput. Chem.*, **17**, 19 (1996).
28. M. Orozco and F. J. Luque, *J. Comput. Chem.*, **14**, 587 (1993).
29. C. Alhambra, F. J. Luque, and M. Orozco, *J. Phys. Chem.*, **99**, 3084 (1995).
30. M. Orozco and F. J. Luque, *Theoretical and Computational Chemistry*, Vol. 3, J. S. Murray and K. Sen, Eds., Elsevier, New York, 1996, p. 181.
31. M. M. Francl, *J. Phys. Chem.*, **89**, 428 (1985).

32. R. Bonaccorsi, R. Cimiraglia, P. Palla, and J. Tomasi, *Int. J. Quantum Chem.*, **24**, 307 (1983).
33. K. Kitaura and M. Morokuma, *Int. J. Quantum Chem.*, **10**, 325 (1976).
34. See: J. Tomasi, B. Mennucci, and R. Cammi, *Theoretical and Computational Chemistry, Vol. 3*, J. S. Murray and K. Sen, Eds., Elsevier, New York, 1996, p. 1.
35. M. M. Francl, MEPHISTO program, *QCPE Bull.*, **5**, 30 (1985).
36. M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. A. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Anfres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defress, J. Baker, J. J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople, *Gaussian-94, Rev. A.1*, Gaussian, Inc., Pittsburgh, PA, 1995.
37. F. J. Luque and M. Orozco, Unpublished version of MOPETE computer programs, University of Barcelona, Barcelona, Spain, 1997.
38. W. L. Jorgensen, *BOSS, Version 3.4*, Yale University, New Haven, CT, 1990.
39. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. Impey, and M. J. Klein, *Chem. Phys.*, **79**, 296 (1983).
40. P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, **28**, 213 (1973).
41. F. A. Momany, *J. Phys. Chem.*, **82**, 592 (1978).
42. U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, **5**, 129 (1984).
43. M. Orozco and F. J. J. Luque, *Comput.-Aided Mol. Des.*, **4**, 411 (1990).
44. W. J. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.*, **51**, 2657 (1969).
45. J. S. Binkley, J. A. Pople, and W. J. Hehre, *J. Am. Chem. Soc.*, **102**, 939 (1980).
46. A. D. McLean and G. S. Chandler, *J. Chem. Phys.*, **72**, 5639 (1980).
47. M. J. Frisch, J. A. Pople, and J. S. Binkley, *J. Chem. Phys.*, **80**, 3265 (1984).
48. T. Clark, J. Chandrasekhar, G. W. Spitznagel, and P. v. R. Schleyer, *J. Comput. Chem.*, **4**, 294 (1983).
49. J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, **117**, 4177 (1995).
50. M. J. Frisch, M. Head-Gordon, and J. A. Pople, *Chem. Phys. Lett.*, **166**, 281 (1990).
51. S. R. Gadre, I. H. Shrivastava, and S. A. Kulkarni, *Chem. Phys. Lett.*, **170**, 271 (1990).